Lab03\_Massey

Matt Massey

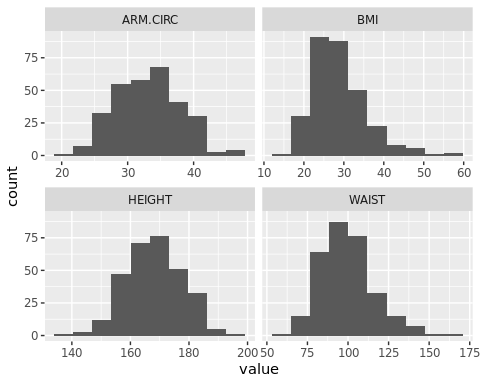
6/4/2021

**Part 1. Assessing Normality**

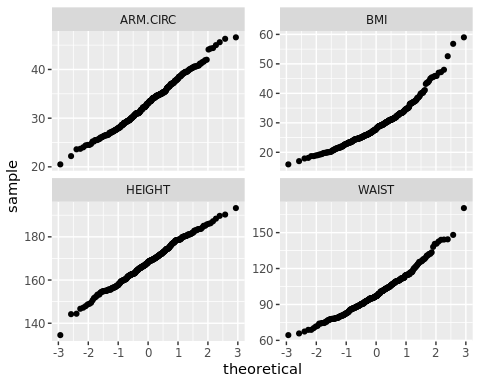
#The library function loads or "wakes up" the packages.   
library(tidyr)  
library(ggplot2)  
  
#Next, import the data set. See video for details.  
data <- read.table(file = "bodydata.txt", sep = "\t", header = TRUE)  
  
# With the tidyr and ggplot2 packages, we can simultaneously look at several of the histograms and normal quantile plots. We will only look at three variables at a time. The code below is only for Height, Waist, Arm circumerference, and BMI. This code is a little complex. Make sure you copy it exactly as you see it.   
data[,12:15] %>% gather() %>% head()

## key value  
## 1 HEIGHT 172.0  
## 2 HEIGHT 186.0  
## 3 HEIGHT 154.4  
## 4 HEIGHT 160.5  
## 5 HEIGHT 179.0  
## 6 HEIGHT 166.7

ggplot(gather(data[,12:15]), aes(value)) +   
 geom\_histogram(bins = 10) +   
 facet\_wrap(~key, scales = 'free\_x')



#Similarly, we can look at the normal quantile plots for each column. The quantile plots are also called QQ plots.   
ggplot(gather(data[,12:15]), aes(sample=value)) +   
 stat\_qq() +   
 facet\_wrap(~key, scales = 'free\_y')



Based on histograms and the quantile plots, height and arm circumference are good approximations of normal distributions - both histograms are symmetric and the quantile plots approximate a straight line. The quantile plot of arm circumference shows a slight deviation of the largest data points, but I would still approximate this as a normal distribution.

In contrast, both BMI and waist are not normally distributed. Histograms of both variables show positive skews to the right (largest values). The quantile plots also show deviations from a straight line and a concave-up trend, although this is more obvious in BMI.

**Part 2. Binomial Probabilities**

#We can use the following information to approximate a normal distribution  
#mu = np, where n is the number of subjects, and p is probability  
#sigma = square root of npq, where q = 1-p  
  
#Problem 9: Find the probability that at least 40 of the 100 subjects have blue eyes.   
x = 40  
n = 100  
p = .35  
pbinom(x, size = n, prob = p, lower.tail = FALSE, log.p = FALSE)

## [1] 0.1250228

#The Normal approximation code is   
pnorm(x, mean = n\*p, sd = sqrt(n\*p\*(1-p)), lower.tail = FALSE)

## [1] 0.1472537

#Problem 10: Find the probability that at least 49 of the 100 subjects have blue eyes.   
x = 49  
n = 100  
p = .35  
pbinom(x, size = n, prob = p, lower.tail = FALSE, log.p = FALSE)

## [1] 0.001450561

#The Normal approximation code is   
pnorm(x, mean = n\*p, sd = sqrt(n\*p\*(1-p)), lower.tail = FALSE)

## [1] 0.001666677

#Problem 11: Find the probability that fewer than 5 of the 100 subjects have green eyes, with p = .12  
#We use x = 4 for binomial, since it is fewer than 5  
x = 4  
n = 100  
p = .12  
pbinom(x, size = n, prob = p, lower.tail = TRUE, log.p = FALSE)

## [1] 0.005256594

#The Normal approximation code is  
pnorm(x, mean = n\*p, sd = sqrt(n\*p\*(1-p)), lower.tail = TRUE)

## [1] 0.006911512

#Question 12: Find the probability that 33 or fewer of the 100 subjects have brown eyes, with p = .4  
x = 33  
n = 100  
p = .4  
pbinom(x, size = n, prob = p, lower.tail = TRUE, log.p = FALSE)

## [1] 0.0912536

#The Normal approximation code is   
pnorm(x, mean = n\*p, sd = sqrt(n\*p\*(1-p)), lower.tail = TRUE)

## [1] 0.07652094

Question 9.

The binomial probability is 0.1250228 and the normal approximation is 0.1472537. Probability calculated from the normal distribution is close and is a good approximation.

Question 10.

The binomial probability is 0.001450561 and the normal approximation is 0.001666677. Probability calculated from the normal distribution is close and is a good approximation.

Question 11.

The binomial probability is 0.005256594 and the normal approximation is 0.006911512. Probability calculated from the normal distribution is close and is a good approximation.

Question 12.

The binomial probability is 0.0912536 and the normal approximation is 0.07652094. Probability calculated from the normal distribution is close and is a good approximation.